Appendix

To avoid inaccurate standard errors associated with hospital and patient effects and to determine how much of the variability in outcomes can be explained by individual- and hospital-level factors (Slonim et al. 2007), we also use hierarchical (multilevel) analysis to test our hypotheses.

In the hierarchical model, we test the explanatory power of variation in group-level variables by examining intraclass correlations (ICC[1])(McGraw;, and Wong 1996). We also describe the effect using a median odds ratio (MOR) with a 95% credible interval (95% CrI). The MOR shows that when two people from different hospitals are randomly selected, variation among hospitals increases an individual's odds of not being enrolled in a P4P program in the median case (Merlo et al. 2006). We calculate the standard errors for hospital-level variance to obtain the 95% CrI for the MOR (Schootman et al. 2007). We test the model in three steps. First, we estimate a null model that has no predictor at either level 1 (the patient level) or level 2 (the hospital level) to divide the measure of variance into within- and between-hospital components. Second, to obtain a precise estimation, patient measures are regressed on the grand mean-centered, individual-level predictors (random intercept model), therefore reducing multicollinearity (David, and Mark 1998). In the third step, the level 2 analysis, we use the intercept estimates obtained from the level 1 as outcome

variables and regress these on the hospital-level predictors to assess the main effects of hospital-level factors (intercept-as-outcome model). We also compute the proportion of the total variance in patient measure explained by hospital-level factors (R^2 between-hospitals) (Raudenbush, and Bryk 2002), and subject the difference in variance to a chi-squared test to determine whether the new model achieved a better fit (Cho 2003).

In the random intercept model, the ICC(1) is .65, indicating that 65 percent of the variance is contributed by hospital factors and 35 percent is contributed by patient factors (Table 1). The MOR is 10.56 (2.92, 23.48). It seems that hospital characteristics contribute largely to the exclusion of DM patients from P4P programs. After adjusting for individual characteristics at the patient level, group predictors account for 27.99% of the initial hospital-level variance, (6.11-4.40)/6.11. Compared with the null model, the random intercept model produces a better fit, with a significant difference of deviance of 1471970.2 ($p < 0.001$). The intercept-as-outcome model also has a deviance statistic that is fairly different from that of the random intercept model and with a very significant chi-squared value ($p < 0.001$). This model also shows a better fit than the random intercept model. From these results, it is clear that there is also patients with more comorbidities or more severe conditions are prone to be excluded from P4P programs.

The mixed-effects model was robust from two perspectives. The results were reliable because the average number of DM patients treated by hospitals was greater than five (multilevel models must use an average of only five observations per group) (Clarke 2008). In addition, select multilevel paradigms have significant implications in the choice of centering. In this study we proposed the use of grand mean centering instead of raw metric methods to reduce potential problems of multicollinearity (David, and Mark 1998). Although the (standard error)/(coefficient) ratios of individual variables from the mixed-effects model tend to be higher compared with those of the logistic model (Bardenheier et al. 2005; D'Errigo et al. 2007), in this study the mixed-effects model does not much alter the magnitude of effects. In other studies, we also find that the odds ratios of individual variables of mixed-effects and logistic regression estimates were slightly different (Bardenheier et al. 2005; D'Errigo et al. 2007).

Table 1 Factors associated with exclusion of DM patients from P4P programs

	logistic model	Mixed-effects model	
		Random intercept	Intercept-as-outcome
	OR (95%CI)	OR (95%CI),τ ※	OR (95%CI),τ ※
Patient level			
Intercept	5.27 (5.23,5.30)***	1.41(1.27,1.56)*** ,6.11*** ※	1.07 (1.05,1.10)*** ,4.40*** ※
Age (Ref: <=63)			
>63	1.32 (1.30,1.33)***	1.20 (1.17,1.23)***	1.05 (1.04,1.06)***
Gender (Ref: F)			
Male	1.11 (1.09,1.12)***	1.06 (1.05,1.08)***	1.05 (1.04,1.06)***
DCSI score(Ref: 0)			
1	0.81 (0.79,0.82)***	0.89 (0.85,0.96)***	0.88 (0.86,0.91)***
2	1.31 (1.29,1.34)***	1.30 (1.24,1.37)***	1.07 (1.04,1.11)***
3	1.12 (1.08,1.15)***	1.17 (1.11,1.23)***	1.05 (1.00,1.10)*
4	1.70 (1.62,1.79)***	1.62 (1.52,1.72)***	1.24 (1.17,1.32)***
5+	1.78 (1.64,1.92)***	1.71 (1.60,1.82)***	1.44 (1.35,1.53)***
CIC count (Ref: 0)			
1	1.08 (1.06,1.10)***	1.10 (1.08,1.13)***	1.05 (1.03,1.07)***
2	1.21 (1.18,1.24)***	1.26 (1.22,1.31)***	1.15 (1.12,1.18)***
3	1.12 (1.08,1.15)***	1.38 (1.32,1.45)***	1.19 (1.15,1.23)***
4+	1.33 (1.24,1.42)***	1.51 (1.40,1.62)***	1.35 (1.28,1.43)***
Number of visits	0.96 (0.95,0.96)***	0.96 (0.96,0.97)***	0.92 (0.92,0.93)***
Hospital level			
Level (Ref: clinic)			
Tertiary hospital	8.13 (7.90,8.36)***		3.44 (2.01,5.87)**
Regional hospital	4.02 (3.94,4.10)***		2.29 (1.82,2.89)***
District hospital	2.98 (2.92,3.05)***		1.86 (1.54,2.25)***
Baseline score#	0.98 (0.98,0.98)***		0.98 (0.98,0.99)***
Patient volume	1.00 (1.00,1.00)*		1.00 (1.00,1.00)
C index	0.72		
ICC(1)		0.65	0.57
MOR(95CrI)		10.56(2.92,23.48)	7.40(1.67,16.18)
$R^2_{\text{between-hospitals}}$			27.99%
Deviance (χ^2test)+		1471970.2*** +	1465002.4*** +

Note: Ref=reference group; *p<0.05 ; **p<0.01 ; ***p<0.001;

#: Prior year; ※: The estimations of the random variance components (τ_s) are in parentheses. The τ_s for the intercepts also represent the between-hospital variance in

patients being enrolled in P4P program. $^{+}$: compared to the previous models, respectively (null model or random intercept model)

Patient all-cause mortality was identified by the records of in-hospital deaths in the claim data (Chen, Ho, and Li 2006). A total of 2115 patients (2115/884452=0.2%) were categorized under all-cause inpatient mortality. We subtracted these patients from our study group. Table 2 in the Appendix shows that the exclusion of these deaths only affected the results slightly in the logistic model.

Table 2. Factors associated with the exclusion of DM patients from P4P programs using a logistic model

	Deleting death	Keeping death
	OR (95%CI)	OR (95%CI)
Patient level		
Intercept	2.77(2.75,2.77)***	5.27 (5.23,5.30)***
Age (Ref: <=63)		
>63	1.29 (1.29,1.29)***	1.32 (1.30,1.33)***
Gender (Ref: F)		
Male	1.11 (1.11,1.11)***	1.11 (1.09,1.12)***
DCSI score(Ref: 0)		
1	0.83 (0.83,0.83)***	0.81 (0.79,0.82)***
2	1.34 (1.34,1.34)***	1.31 (1.29,1.34)***
3	1.10 (1.09,1.10)***	1.12 (1.08,1.15)***
4	1.59 (1.58,1.59)***	1.70 (1.62,1.79)***
5+	1.50 (1.49,1.51)***	1.78 (1.64,1.92)***
CIC count (Ref: 0)		
1	1.05 (1.04,1.05)***	1.08 (1.06,1.10)***
2	1.12 (1.12,1.13)***	1.21 (1.18,1.24)***
3	1.22 (1.21,1.22)***	1.12 (1.08,1.15)***
4+	1.14 (1.14,1.15)***	1.33 (1.24,1.42)***
Number of visits	0.99 (0.99,0.99)***	0.96 (0.95,0.96)***
Hospital level		
Level (Ref: clinic)		
Tertiary hospital	7.75 (7.75,7.75)***	8.13 (7.90,8.36)***
Regional hospital	3.98 (3.97,3.98)***	4.02 (3.94,4.10)***
District hospital	2.59 (2.58,2.60)***	2.98 (2.92,3.05)***
Baseline score[#]	0.99 (0.99,0.99)***	0.98 (0.98,0.98)***
Patient volume	1.00 (1.00,1.00)***	1.00 (1.00,1.00)*
C index	0.71	0.72

Note: Ref=reference group; *p<0.05；**p<0.01；***p<0.001;

[#]: Prior year;